13004–13015 *Nucleic Acids Research, 2017, Vol. 45, No. 22* *Published online 23 October 2017*
doi: 10.1093/nar/gkx982

Multiple RNA structures affect translation initiation and UGA redefinition efficiency during synthesis of selenoprotein P

Marco Mariotti[1], Sumangala Shetty[2], Lisa Baird[3], Sen Wu[4], Gary Loughran[5], Paul R. Copeland[2], John F. Atkins[3,5] and Michael T. Howard[3,*]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA USA, [2]Department of Biochemistry and Molecular Biology, Rutgers Robert Wood Johnson Medical School, Piscataway, NJ USA, [3]Department of Human Genetics, University of Utah, Salt Lake City, UT, USA, [4]State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, People's Republic of China and [5]Biochemistry and Cell Biology, University College Cork, Cork, Ireland

Received August 22, 2017; Revised October 03, 2017; Editorial Decision October 08, 2017; Accepted October 11, 2017

ABSTRACT

Gene-specific expansion of the genetic code allows for UGA codons to specify the amino acid selenocysteine (Sec). A striking example of UGA redefinition occurs during translation of the mRNA coding for the selenium transport protein, selenoprotein P (SELENOP), which in vertebrates may contain up to 22 in-frame UGA codons. Sec incorporation at the first and downstream UGA codons occurs with variable efficiencies to control synthesis of full-length and truncated SELENOP isoforms. To address how the *Selenop* mRNA can direct dynamic codon redefinition in different regions of the same mRNA, we undertook a comprehensive search for phylogenetically conserved RNA structures and examined the function of these structures using cell-based assays, *in vitro* translation systems, and *in vivo* ribosome profiling of liver tissue from mice carrying genomic deletions of 3′ UTR selenocysteine-insertion-sequences (SECIS1 and SECIS2). The data support a novel RNA structure near the start codon that impacts translation initiation, structures located adjacent to UGA codons, additional coding sequence regions necessary for efficient production of full-length SELENOP, and distinct roles for SECIS1 and SECIS2 at UGA codons. Our results uncover a remarkable diversity of RNA elements conducting multiple occurrences of UGA redefinition to control the synthesis of full-length and truncated SELENOP isoforms.

INTRODUCTION

Selenoproteins are a class of proteins that contain the amino acid selenocysteine (Sec) (1). Many selenoproteins have oxido-reductase activities that strictly depend on the Sec residue located at the active site to provide catalytic redox function. Biosynthesis of selenoproteins is unique in that the incorporation of Sec occurs during translation in response to an in-frame UGA codon, which in standard decoding specifies termination. To achieve such programmed 'recoding' of the genetic code requires the coordinated action of several *trans*- and *cis*-acting factors (2).

In eukaryotes and archaea UGA redefinition to Sec depends on the presence of *cis*-acting Sec insertion sequences (SECIS) located in the 3′ UTR (3–6). In contrast, bacterial SECIS elements are located a few nucleotides 3′ of the UGA codon within the open reading frame (7,8). Some eukaryotic selenoprotein mRNAs also contain a *cis*-acting structure adjacent to the UGA, known as the Sec redefinition element (SRE) (9–12). However, unlike the SECIS elements, SREs are not present in all selenoprotein genes and are not essential for Sec incorporation but do enhance recoding activity. During translation, selenoprotein transcripts recruit the SECIS RNA binding protein (SECISBP2) (13,14) and a specialized Sec elongation factor (eEFSec) that delivers Sec-tRNA$^{[Ser]Sec}$ to the ribosome (15–17). In addition to these core components of the Sec-incorporation machinery are other *trans*-acting factors, such as eIF4a3 (18), L30 (19), and nucleolin (20) that have been reported to regulate gene-specific Sec incorporation efficiencies.

The process of UGA redefinition and Sec incorporation has consistently been shown to be inefficient due, at least in part, to the requirement for specialized Sec insertion factors to recruit the Sec-tRNA$^{[Ser]Sec}$ in a process that competes with termination of translation. As most selenopro-

*To whom correspondence should be addressed. Tel: +1 801 585 1927; Email: mhoward@genetics.utah.edu

tein mRNAs contain only a single UGA codon, inefficient Sec incorporation may be tolerated, however, a remarkable exception in the one-mRNA, one-Sec paradigm is found in the vertebrate *Selenop* gene (previously known as Sepp1 or SelP) (21). Human and rodent *Selenop* mRNAs each have 10 UGAs (5) and in other vertebrate species the UGA count may range from 7 to 22 depending on the species (22,23) (Mariotti M., unpublished). The products of *Selenop* translation consist of a shortened N-terminal isoform having the first Sec residue in a thioredoxin-like motif with presumed peroxidase activity (24) terminating at the second Sec residue, as well as longer isoforms that terminate at Sec positions within the C-terminal Sec-rich domain (25,26) or at the natural termination codon. Whereas the biological role of the peroxidase activity is unknown, the longer isoforms are critical for selenium delivery to the brain, testes, and other tissues (27) and may have other as of yet uncharacterized functions. Despite the aforementioned inefficiency in Sec incorporation, the production of long SELENOP isoforms demands that Sec incorporation at some UGA-Sec codons must occur with high efficiency while production of the short isoform requires some UGA codons to terminate translation suggesting that unique mechanisms may be involved in translation of the *Selenop* mRNA.

Studies of *Selenop* translation using a SECISBP2 supplemented *in vitro* translation system indicated that after incorporation of selenocysteine at a first UGA codon, incorporation at downstream UGA codons occurs with much higher efficiency (28,29). These studies suggest that successful recoding of the first UGA codon shifts downstream UGA decoding in favour of Sec incorporation. *Selenop* mRNAs are also unique in containing two 3′ UTR SECIS elements with other vertebrate selenoproteins containing only a single SECIS element. Cell culture experiments from the laboratory of Marla Berry have indicated that each SECIS element may have distinct functions in decoding mRNAs with multiple UGA codons (30). This finding led the authors to propose an innovative spatially restricted model in which each SECIS element contributes to the regulation of Sec incorporation but with SECIS2 mediating redefinition of the first UGA and SECIS1 mediating redefinition of the remaining UGAs residing near the 3′ end of the coding sequence. However, the mechanism by which two SECIS elements located in the 3′ UTR act to reprogram the ribosome with differing efficiencies in different regions of the *Selenop* mRNA, and whether other additional RNA structures may be involved to regulate *Selenop* recoding, is not known.

Our understanding of the mechanism by which the ribosome is reprogrammed at multiple sites in the same RNA during translation to incorporate Sec has been limited by the inability to quantify the formation of unstable termination products (31) relative to the various full-length and near full-length SELENOP proteins *in vivo*. To begin to address this problem, we recently described SELENOP isoforms present in serum from mice carrying genomic deletions of either *Selenop* SECIS1 or SECIS2 (32). Of direct biological relevance to this study is the demonstration that in primates 10–25% of *Selenop* mRNA naturally lacks SECIS 2 due to the utilization of a cryptic poly-(A) site located between SECIS1 and SECIS2 (32). The analysis of protein products produced from the SECIS deleted mice support that the deletion of SECIS1 affects the levels of the long SELENOP isoform and deletion of SECIS2 reduces overall SELENOP abundance (32), however, the presence of multiple isoforms, uncertainties about the effects of protein processing and glycosylation on migration patterns, and the inability to identify truncated products at the first UGA limited mechanistic interpretations of the data.

To further investigate how the *Selenop* mRNA directs multiple instances of UGA redefinition, we performed a comprehensive search for conserved RNA structures in vertebrate *Selenop* genes. Our analysis revealed that, in addition to the two SECIS elements in the 3′ UTR, *Selenop* contains several distinct conserved structures within the coding sequence as well as an increased probability of nucleotide pairing near the beginning and end of the coding sequence. These structures include several previously undescribed SREs and a novel stem loop structure overlapping the signal peptide sequence near the initiation codon. We address the function of these elements using cell based assays and a newly developed and improved dual reporter system that involves StopGo release of the upstream and downstream reporters during translational elongation (33). Furthermore, we characterized the distinct functions of two *Selenop* SECIS elements for the first time *in vivo* by applying ribosome profiling to measure UGA recoding efficiencies in mouse mutants carrying genomic deletions of either SECIS1 or SECIS2.

MATERIAL AND METHODS

Identification of Selenop RNA structures

The database of vertebrate NCBI RefSeq transcripts was searched online with blastn using the canonical mRNA *Selenop* isoform (NM_009155.4) as query. The matching transcripts were aligned and manually inspected to remove partial sequences. The alignment was further processed to contain only a single sequence per species, the one most similar to the mouse canonical isoform. This resulted in a final set of 131 mRNA sequences (Supplementary Material S1 and S2), which was searched using RNAz v2.1 (34). The algorithm of RNAz can process at most 6 sequences. To avoid biasing our analysis toward any particular species in the alignment, we ran 200 iterations in which six species were selected at random from the alignment, and their sequences were fed to RNAz. The result was a set of RNA-class probabilities assigned to each alignment window. Three window sizes were considered (60, 90, 120 nucleotides; step size = 10) to make sure no RNA element was missed because of its length. Since the identification of RNA elements depended on the correctness of the input nucleotide alignment, we tested various programs and parameters, checking which one led to the identification of SECIS elements in the 3′UTR. Our final choice was MAFFT v.7215 with default parameters (35). In order to produce a visual representation of the RNA structures, we folded sequences using RNAalifold (36). Since using the full alignment did not give satisfactory results, instead we ran RNAalifold providing the 20 sequences appearing most often for each element in RNAz outputs with probability >0.95 (after removing alignment columns with >40% gaps). The precise boundaries of the RNA elements were determined empirically: we

ran the same procedure with varying starting and ending positions, checking whether the predicted structure was extended or remained the same.

Cell transfection and in vitro translation-based reporter assays

The HuSELENOP cDNA was cloned into pcDNA3.1 vector. The HuCysSELENOP, the coding region of the artificial selenoprotein (Luc10UGA) and (Luc10UGC) as well as the N-terminal and C-terminal chimeras HuN-term_Luc and Luc_HuC-term were synthesized commercially by Integrated DNA technologies (IDT). These constructs were then cloned into pcDNA3.1 vector and the HuSELENOP 3′UTR was then ligated into the Sec containing constructs using Pac1 and Not1 linkers. The resulting sequences were confirmed by DNA sequence analysis.

In vitro translation assay was performed as described earlier (29). Briefly, *in vitro* transcribed and capped mRNA of either human *Selenop* wild type or mutants were translated in rabbit reticulocyte lysate supplemented with 8 pmol of CT-SBP2 and either [^{35}S]Cys or [^{75}Se]. Translation reactions were incubated for 1 h at 30°C. 4μl of the translation products were then resolved by 12% SDS-PAGE gel and quantitated by PhosphorImager analysis (GE Healthcare).

Reporter constructs to examine the ISL structure were made by inserting the two *Selenop* SECIS elements (contained within nts 1461–2030 (NM_009155)) into the XbaI site of phRL-CMV (Addgene) downstream from the *Renilla* luciferase coding sequence. Next, the first 210 nts of *Selenop* (NM_009155) were inserted into the NheI and AvaII sites of phRL-CMV (Addgene) upstream and in-frame with the *Renilla* luciferase coding sequence. The U40S mutation replaces the Sec encoding TGA with TCA. ISLdis constructs had nucleotides 107, 110, 113, 116, 140, 143, 146, 149, 158, 167 and 170 (NM_009155) changed to the complementary nucleotide. These positions were selected to maintain the amino acid encoded while disrupting the predicted RNA structure. *In vitro* translations were conducted in the presence of L-[35]-Methionine using the TnT T7 Quick Coupled Transcription /Translation System (Promega) in the presence or absence of exogenous SECISBP2 protein.

Reporter constructs to analyze SRE1 were similarly designed by inserting the two SECIS elements into the XbaI site of pSGDluc (33). The SECIS1-only and SECIS2-only constructs contained nts 1461–1840 and nts 1625 to 2030 (NM_009155) inserted into the Xba1 site, respectively. Nucleotides 210–317 (NM_009155) containing the first UGA and SRE1 were then inserted between the PspX1 and BglII sites such that the intervening coding sequence was in-frame with the upstream *Renilla* and downstream firefly luciferase coding sequence. Plasmids were transfected into HEK293 cells in $\frac{1}{2}$ area 96 well tissue culture treated cells using Lipofectamine 2000 (Thermo Fisher Scientific). Lipofectamine:DNA complexes were formed by mixing 25 ng of DNA + 12.5 ul of Opti-MEM (Gibco) with 0.2 ul of Lipofectamine 2000 + 12.5 ul of Opti-MEM (Gibco). After incubation for 20 min at RT, 25 ul of solution is placed in each well and overlaid with 50 ul of HEK293 cells (8×10^5 cells/ml) in DMEM (Gibco) + 10% FBS (HyClone). Cells are incubated overnight at 37°C incubator with 5% CO_2 and 75 ul of DMEM + 10% FBS. Firefly and *Renilla* luciferase activities were determined using the Dual Luciferase Stop & Glo® Reporter Assay System (Promega). Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems). Transfected cells were lysed in 12.5 μl of 1 × passive lysis buffer (PLB) and light emission was measured following injection of 25 μl of either *Renilla* or firefly luciferase substrate. Redefinition efficiencies (% Recoding) were calculated as the ratio of firefly activity/*Renilla* activity for the test (UGA) sequences as a percentage of the ratio of Firefly activity/*Renilla* activity for the corresponding UCA control sequence. Mean and standard deviations were calculated based on at least six independent transfections.

Mice

Mice used in this study consisted of C57BL/6 (wild type) and two SECIS deletions strains, *Sepp1*$^{\Delta SECIS1}$ and *Sepp1*$^{\Delta SECIS2}$, previously described (32). Whole livers were excised from male mice ($n = 3$) and rapidly frozen in liquid nitrogen.

Ribosome profiling and RNA-Seq

For ribosome profiling, ∼100 mg of material was suspended in 1.5 ml of lysis buffer (10 mM Tris–Cl (pH 7.5), 300 mM KCl, 10 mM $MgCl_2$, 200 ug/ml cycloheximide, 1 mM DTT and 1% Triton X-100 and homogenized in the Mini-Beadbeater-8 (Biospec Products) with 3 × 2.3 mm diameter chrome steel balls for 2 × 30 s. Insoluble debris was removed by centrifugation at 12 000 × g at 4°C. Next 600 U of RNase1 (Ambion) were added and the sample was incubated at RT for 45 min. Ribosomes were isolated by centrifugation through 50% sucrose at 200 000 × g for 3.5 h. Ribosomes pellets were resuspended in Qiazol (Qiagen) and ribosome protected fragments were isolated using the miRNAeasy kit (Qiagen) with modification to retain small RNAs, as described by the manufacturer. For RNA-Seq, ∼60 mg of tissue was homogenized as described above with the exception that the lysis buffer was replaced with 1.5 ml of Qiazol and subsequently purified using the miRNAeasy Mini Kit (Qiagen) as described by the manufacturer. Poly(A) mRNA was isolated using the Poly(A) Purist Mag kit (Ambion) and randomly fragmented by heating in the presence of magnesium. rRNA was removed from both ribosome profiling and RNA-Seq samples using the Ribo-Zero Gold rRNA Removal Kit (Illumina) as described by the manufacturer. Both RPFs and randomly fragmented PolyA enriched total RNA were electrophoresed on a 15% TBE Urea gel and RNA fragments between 20 and 40 nts in size were purified prior to the construction of libraries for deep-sequencing. Small RNA sequencing libraries were constructed using the Illumina TruSeq Small RNA Sample Prep kit (Illumina), according to the manufacturer's instructions. Libraries were subjected to 50-cycle single-end sequencing on the Illumina HiSeq 2000 Instrument. Data can be obtained from the NCBI GEO repository, entry TBD. Adapter sequences were trimmed from all sequences using the FASTX-Toolkit from the Hannon Lab (hannonlab.cshl.edu/fastx_toolkit/). Contaminating rRNA sequences were removed using bowtie

(37) by alignment to the mouse 43sRNA repeat sequence and preservation of all unaligned sequences. Uniquely mapping sequences (million mapped reads) were identified by alignments using bowtie to RefSeq mRNA entries obtained from the UCSC browser (mm9) in which all mRNAs derived from the same gene were reduced to a single entry corresponding to the longest isoform. Direct alignments to *Selenop* (NM_009155) were performed and custom perl scripts were used to determine the A-site position and sequence read count across the mRNA.

RESULTS

RNA structures of vertebrate *Selenop*

We set out to identify all *cis*-acting RNA structures in *Selenop*. In order to detect biologically relevant elements, we took a comparative approach aimed to discover RNA structures conserved during evolution. We manually curated an alignment of 131 *Selenop* mRNA sequences across vertebrates (Supplementary Material S1) for analysis with RNAz (38). RNAz is designed to identify structures supported by multiple sequences requiring evidence of compensatory variation in sequence to support the predicted structures. RNAz scans the input alignment in sliding windows, and can process only six sequences at a time. To overcome this limitation, we ran RNAz with three different window sizes, and performed extensive random sampling of sequences from our alignment. For each sample and at each sequence window, we recorded the SVM RNA-class probability output by RNAz, which peaks in presence of conserved structures. Results are shown in Figure 1. Our procedure identified the two SECIS elements in the 3′UTR of *Selenop*, and highlighted novel structures located in the coding sequence. Using RNAalifold (36), we proceeded to determine the boundaries of these elements and to predict their structures (Materials and Methods), shown in Figure 2.

The first structure, which we refer to as the Initiation Stem Loop (ISL), was located at the very beginning of the coding sequence, starting just a few codons downstream of the initiator AUG. ISL was the most obvious coding sequence RNA element resulting from our analysis and, according to the free energy predicted by RNAalifold, by far the most stable (Figure 2A). The two previously described SECIS elements were readily identifiable in the 3′ UTR (Figure 2B and C)

The remaining structures were predicted in proximity of Sec encoding UGA codons, and thus were named SRE, after the analogous elements previously identified in other selenoprotein genes (9–12). SRE1 overlaps the first Sec UGA of *Selenop* (Figure 2D). It should be noted that although we are terming the new UGA-linked structures as SREs, they may be functionally distinct from SREs described in other selenoprotein mRNAs. The first Sec residue is conserved in all vertebrate SELENOP proteins. It is located at the active site of the SELENOP N-terminal domain, which has a thioredoxin-like fold responsible for the oxido-reductase activity of SELENOP (24). The rest of the Sec residues in SELENOP are found in a Sec-rich C-terminal domain. The different number of Sec residues among different species is due to variations in this region. Homologous Sec positions

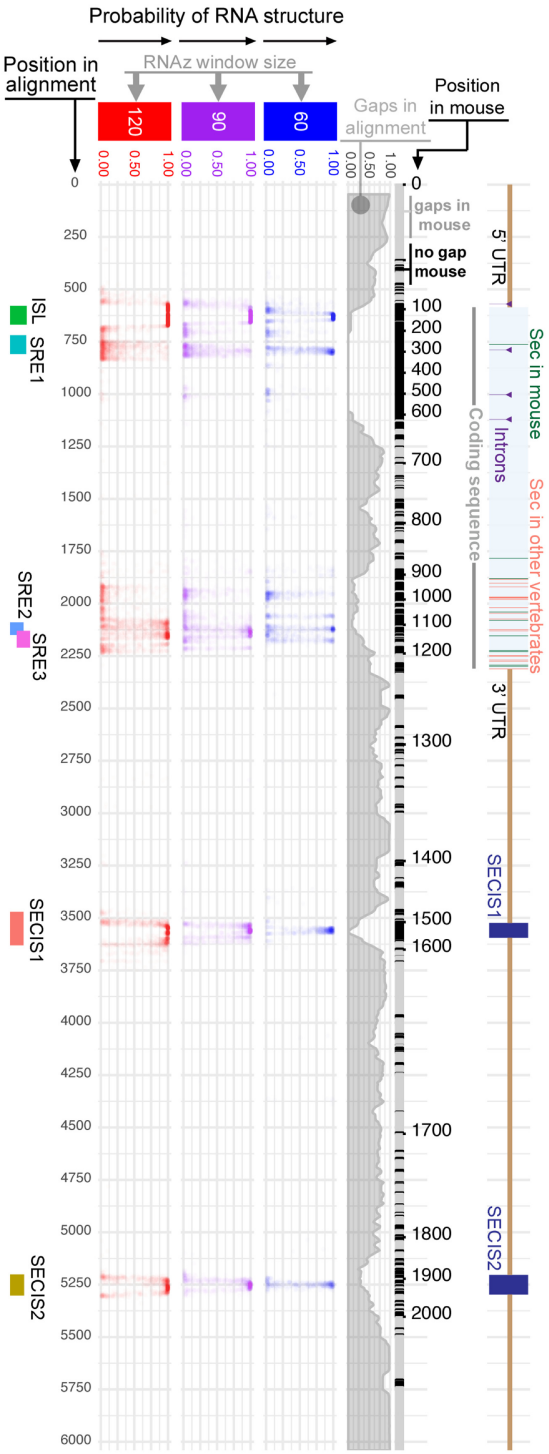

Figure 1. Identification of RNA structures in *Selenop* conserved across vertebrates. An alignment of 131 vertebrate mRNAs was searched for conserved structures using RNAz. The plot shows the RNAz probability at each alignment position, for three different window sizes. Only the probabilities greater than 1% are plotted. Various gene features (coding sequence, Sec residues, intron boundaries, and SECIS elements) were mapped from the mouse sequence to the corresponding alignment position; these are shown on top. The second line serves as a guide to map the positions in the mouse sequence to the alignment. The third line shows the percent of gaps in the alignment computed in sliding windows of 90 nucleotides. At the bottom of the plot, the boundaries of the structures displayed in Figure 2 are shown.

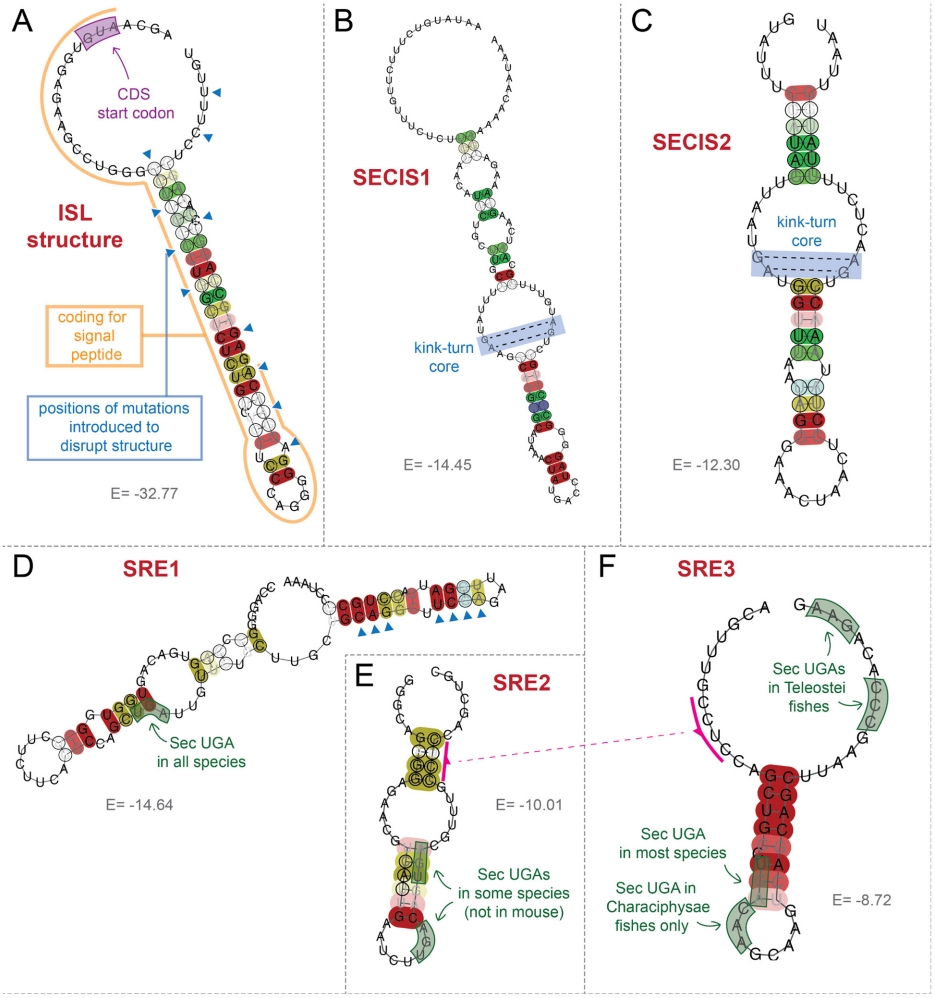

Figure 2. Predicted structures of Selenop RNA elements. RNA structures were predicted using RNAalifold (Materials and Methods). The predicted free energy in kcal/mol is shown for each structure. The regions extracted from the alignment for each element are displayed at the bottom of Figure 1. SRE2 and SRE3 reside in adjacent locations: the magenta mark points to the same sequence portion in the two structures. In panel **B** and **C**, the GA quadruplex pairing of the kink-turn motifs are indicated by dashed lines.

are mostly found carrying either Sec-UGAs or codons coding for cysteine, but this is not always the case (Supplementary Material S2). Our RNAz search clearly highlighted the presence of conserved structures in this multi-Sec region. Their boundaries, though, were difficult to pinpoint, and vary among different species. Nevertheless, we could identify at least two widely supported structures: SRE2, located between the fifth and the sixth UGA codon in the mouse *Selenop* mRNA; and SRE3, located just downstream of SRE2, overlapping the sixth UGA codon.

Both N-terminal and C-terminal coding regions are essential for efficient Sec incorporation *in vitro*

The sequence and structural conservation noted above suggest that both N- and C-terminal coding sequences may be involved in UGA redefinition during synthesis of SELENOP. To directly test the role of coding region sequences, we generated an artificial selenoprotein (Luc10UGA), which consists of a luciferase coding region fragment that contains 10 Sec codons spaced exactly as found in human SELENOP (huSELENOP; Figure 3A). Luc10UGA also contains the human signal peptide, the full human *Selenop* 3′ UTR, and a C-terminal FLAG tag. As a control, we also created the Luc10UGC construct, changing the Sec codons to Cys (Figure 3A). As shown in Figure 3B, *in vitro* translation of capped native huSELENOP mRNA in the presence of [^{75}Se] yielded a distinct ∼47kDa protein (note this is 3 kDa larger than the predicted molecular weight of 44 kDa). In contrast, translation of Luc10UGA produced protein that was visualized as a weak and diffuse band at ∼44kDa that likely represents full length protein as well as early termination products at one or more of the C-terminal UGA codons. A 33 kDa product, which would be predicted to result from termination at the second UGA codon, is also detectable, albeit barely (Figure 3B, lane 3). We repeated *in vitro* translation in the presence of [^{35}S]Cys of the cysteine versions huCysSELENOP and Luc10UGC respectively (Figure 3B, right), which demonstrates that the difference in migration of huSELENOP and Luc10UGA translated products is intrinsic and not due to

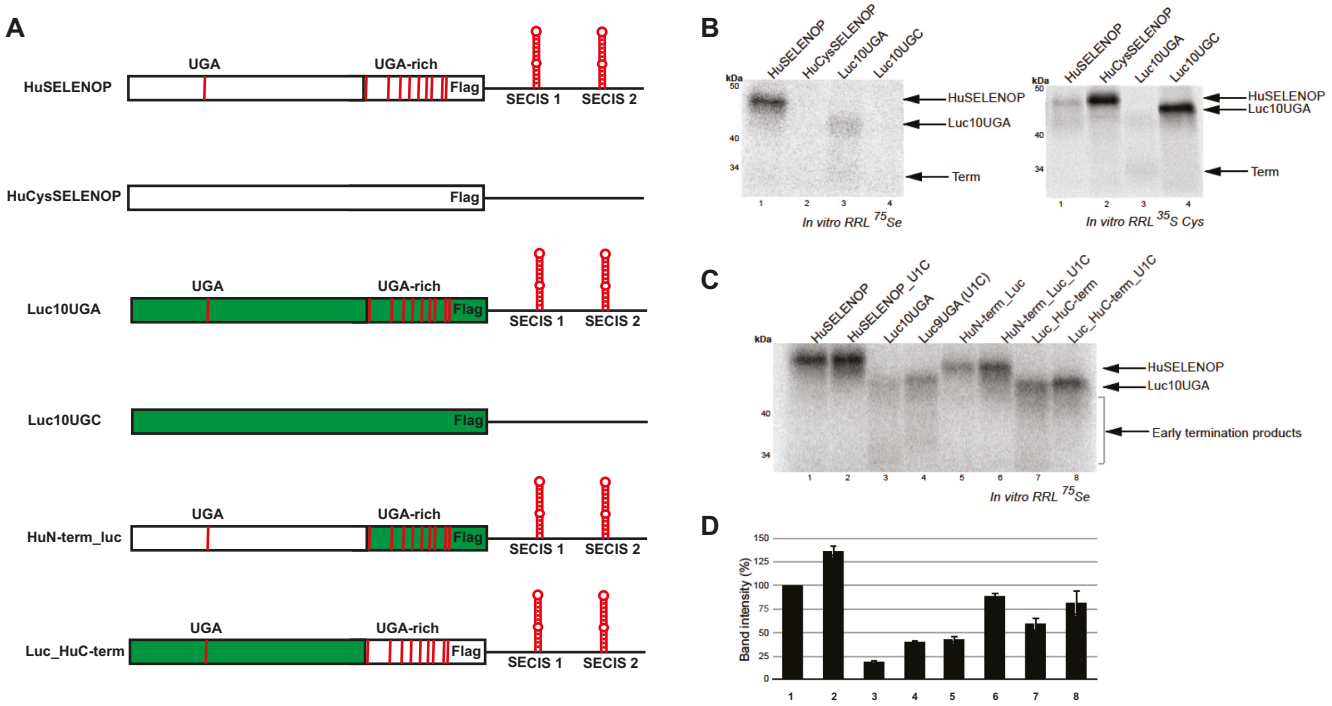

Figure 3. SELENOP N-terminal and C-terminal coding region together are required for efficient production of full-length product. (**A**) Top panel, Schematic diagram of human SELENOP, artificial selenoproteins and chimeras used in this study. White and green boxed areas represent human and firefly sequences, respectively. 3′ UTR sequences are derived from the human *Selenop* gene. (**B**) *Left and Right Panel,* capped huSELENOP, huCysSELENOP, Luc10UGA and Luc10UGC mRNAs were translated in RRL supplemented with CT-SBP2 and analyzed using ^{75}Se and [^{35}S]Cys labeling, respectively. Arrows (from top to bottom) indicate full-length huSELENOP, Luc10UGA and early termination products. Radiolabeled proteins were resolved by SDS-PAGE and detected by PhosphorImager analysis. (**C**) Capped RNA of native huSELENOP and Luc10UGA along with the mutants where the first Sec codon was changed to Cys and the chimeras as shown in panel A were translated in RRL supplemented with CT-SBP2 and analyzed using ^{75}Se-labeling. (**D**) Quantitative analysis of full-length product from the PhosphorImager data in C**.** Data is plotted as the average plus and minus standard deviation for three independent experiments.

premature termination. Together, these results indicate that the native SELENOP coding region sequence is required for efficient and processive Sec incorporation.

To evaluate the contributions made by the RNA sequences coding for the N- and C-termini of SELENOP to processive Sec incorporation, we created the huSELENOP/Luc10UGA chimeras shown in Figure 3A. We fused the sequence coding for the N-terminus of huSELENOP with the C-terminus of Luc10UGA to create HuN-term_Luc, and we conversely fused the mRNA sequences encoding the C-terminus of huSELENOP to those encoding the N-terminus of Luc10UGA just downstream of the first UGA (Luc_HuC-term). Because the first UGA of SELENOP has been earlier implicated to be a 'bottleneck' to processive Sec incorporation (28,30), we also created U1C mutants of these chimeras where the first UGA was mutated to UGC. The N-terminal sequence was defined as the sequence from the start codon up to but not including the 2nd UGA and the C-terminus constitutes sequence immediately downstream from the second UGA to the natural termination codon. mRNAs corresponding to each of the chimeras were translated in RRL in the presence of [^{75}Se]. Interestingly, while all constructs yielded a full-length product, they differed in amounts and the presence of early termination products (Figure 3C). Quantitation of the full-length products normalized to huSELENOP is shown in Figure 3D. Luc10UGA and its U1C mutant both showed a trail of early termination products and the amounts of full-length protein were reduced by 5- and 3-fold respectively. Interestingly, in the presence of either the N- or C-terminus native SELENOP coding sequence, the efficiency increased by 40% and 60% respectively, and the increase in processivity is particularly notable in the cases where the SELENOP N-terminus is present. As expected, all of the U1C mutants displayed increased efficiency, albeit to varying extents. Overall this data indicates that efficient *in vitro* Sec incorporation requires sequences encoding both the N- and C-termini for optimal expression thereby supporting the role for the unique and conserved structures described above.

The ISL promotes translation initiation in an *in vitro* translation system

Immediately downstream of the AUG initiation codon of *Selenop* is the signal peptide required for release of SELENOP into plasma. Overlapping this signal peptide is the ISL described above (Figure 2A). In order to investigate the effects of the ISL on translation of the Selenop RNA, we designed reporter constructs for translation *in vitro* as follows. The 5′ end of mouse *Selenop* was inserted into a *Renilla* luciferase reporter such that the 5′ UTR and the first 210

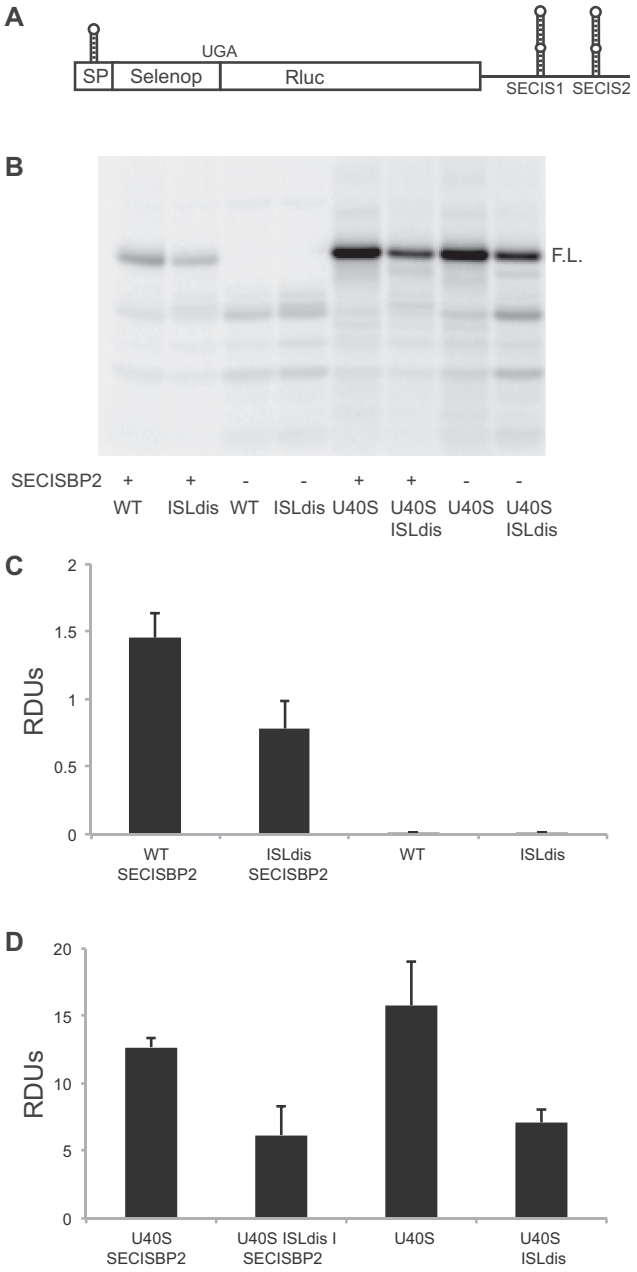

Figure 4. ISL affects translation initiation. Reporter construct represented in (**A**). The thin lines left and right of the boxed region represents 5′ and 3′ UTRs, respectively. ISL, Initiator stem loop; SP, signal peptide; the position of the first UGA in *Selenop* is indicated; Rluc, *Renilla* luciferase; SECIS1 and SECIS2 are located just downstream of Rluc. (**B**) *In vitro* transcription and translation of the indicated reporter constructs in the presence or absence of SECISBP2. 35-S methionine labelled proteins were electrophoresed and visualized by phosphor image analysis. The position of full-length protein is indicated (F.L.). WT, wild type; ISLdis, ISL disruption containing base pair disrupting mutations. U40S, first UGA codon changed to UCA. (**C**) Quantification of full-length protein in relative density units (RDUs) for constructs containing the UGA. (D) Same as C for constructs in which the UGA was changed to UCA.

nts of the coding sequence were in-frame with the reporter. In addition, the two SECIS elements were inserted downstream of the *Renilla* termination codon (Figure 4A). Similarly, we inserted the same region with the first UGA codon (codon 59; amino acid 40 in the mature protein) changed to UCA (U40S). We further produced an ISL variant with mutations that would disrupt the stem loop structure (ISLdis). Given that this region encodes both important amino acids and a potential secondary structure, we altered 11 nts that were located in the third codon position to disrupt RNA structure but maintain the native amino acid sequence (Figure 1A). To avoid potential complications of effects on protein export or RNA stability, reporter constructs were transcribed and translated in rabbit reticulocyte lysates supplemented with [^{35}S] labeled methionine in the presence or absence of SECISBP2 (Figure 4B), which is required for Sec insertion in rabbit reticulocyte lysates (13). Quantification of the radiolabelled products revealed that the ISLdis mutations reduced the production of full-length protein by ∼2-fold irrespective of the inclusion of SECISBP2 or whether there was a UGA or UCA at codon 59 (Figure 4C and D, respectively). These results are consistent with the stem loop increasing the overall rate of translation initiation in reticulocyte lysate translation reactions. Transfection of the same wild type and mutant reporters in cultured cells under conditions of either high or low selenium resulted in highly variable luciferase expression in cell extracts and ineffective transport of the reporter product into the media suggesting a defect in export, folding or instability of the fusion protein in cells (data not shown, see Discussion).

SRE1 affects Sec-insertion at the first UGA in cultured cells

A second RNA structure within the coding region, SRE1, is predicted just downstream of the first UGA in *Selenop* (Figure 2D). We recently developed an improved dual-luciferase reporter plasmid (pSGDluc) to investigate recoding efficiency *in vitro* or in cultured cells (33). The reporter contains the *Renilla* luciferase gene separated from the firefly luciferase gene by a StopGo sequence, a multiple cloning site, and a second StopGo sequence. Insertion of the UGA codon and surrounding sequences (codons 41–76) into the multiple cloning site along with SECIS elements in the 3′ UTR allows recoding efficiency to be monitored as the ratio between *Renilla* and firefly luciferase activities (Figure 5A). The inclusion of StopGo sequences induces release of the nascent chain such that the *Renilla* and firefly luciferase reporters are uncoupled without the addition of variable peptide sequences that could interfere with their activities.

pSGDluc containing wild type codons 41–76 and either SECIS1, SECIS2 or both SECIS elements were transfected into HEK293 cells grown in the presence of low (3 nM) or high levels of sodium selenide (60 nM) and the percent recoding was calculated as described in the materials and methods. To investigate the effect of the downstream SRE1 element on recoding efficiency, mutations were introduced into the lower (M1) or upper (M2) parts of the stem to disrupt base pairing potential (positions indicated in Figure 2D, blue triangles). Base pairing potential in the lower or upper parts of the stem were restored in a second set of mutations directly across the stem and were referred to as R1 and R2, respectively. Mutations in the lower part of the stem reduced recoding by as much as 50%, whereas mutations in the upper part of the stem were reduced by ∼20–30% (Fig-

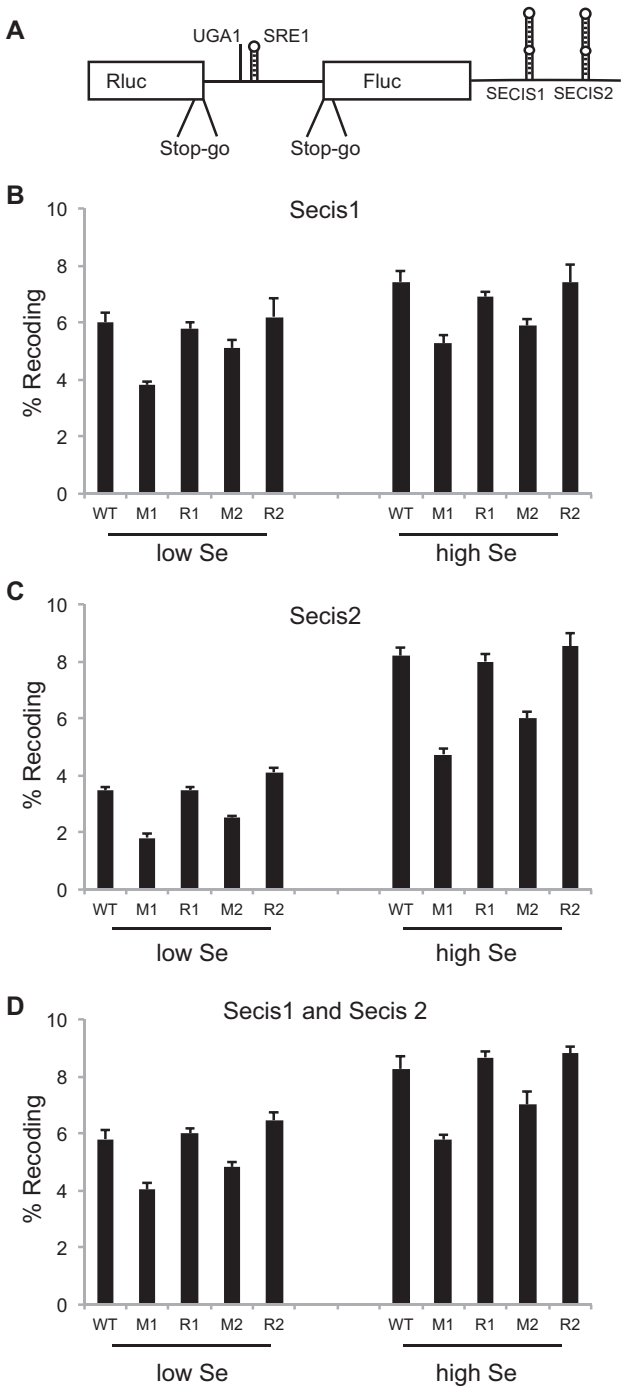

Figure 5. SRE1 affects Sec incorporation at the first UGA codon. Dual luciferase StopGo reporter construct is shown in (**A**). The region of SELENOP containing the first UGA and SRE1 were cloned between the *Renilla* (Rluc) and firefly (Fluc) reporter coding sequences. SECIS1, SECIS2, or SECIS1 and SECIS2 were inserted downstream of the reporter coding sequences to support Sec incorporation. During translation, StopGo sequences located at the 3′ end of Rluc and the 5′ end of Fluc induce peptide cleavage and release of each reporter enzyme respectively. (**B**) Reporter constructs containing SECIS1 were transfected into HEK293 cells and % Recoding is expressed as the ratio of firefly to *Renilla* luciferase normalized to a construct lacking the UGA codon. Mutations M1 and M2 were designed to disrupt the SRE1 structure, whereas R1 and R2 carry compensatory mutations to restore the structure of M1 and M2, respectively. (**C**) Same as B except the reporter contains SECIS2 only. (**D**) Same as B except the reporter contains both SECIS1 and SECIS2.

ure 5B–D). Restoration of base pairing potential returned recoding efficiency to wild type levels. Variations in Se levels had modest effects on constructs containing SECIS1 or SECIS1 and SECIS2, but low levels of Se reduced recoding efficiency by ∼50% in all constructs containing SECIS2 alone suggesting that SECIS2 redefinition activity is sensitive to available selenium levels.

Distinct roles for SECIS1 and SECIS2

To test the function of SECIS1 and SECIS2 *in vivo*, we took advantage of two knockout mouse models, which are deleted for SECIS1 and SECIS2, respectively (32). As SELENOP is produced at high levels in hepatocytes for secretion into plasma, livers from wild type and SECIS deleted mice were excised and analysed by ribosome profiling and RNA-Seq.

As SELENOP is an abundant protein in liver, it is possible that disruption of its synthesis could increase available selenium levels and affect expression of other selenoproteins. RNA-Seq was used to examine the effects of deleting *Selenop* SECIS1 or SECIS2 on other selenoprotein mRNA levels. Sequenced reads mapping to selenoprotein mRNAs were normalized for gene length and total mapped sequences in each sample and expressed as RPKMs (reads per kilobase per million mapped reads) (39) (Supplementary Material S3). We observed that other selenoprotein mRNA levels in liver were unchanged by deletion of SECIS1 and only *Selenop* mRNA was significantly reduced (∼25%) by deletion of SECIS2.

Ribosome profiling involves isolation of ribosome:mRNA complexes from cells or tissues, digestion of the unprotected mRNA with ribonucleases, and deep sequencing of the ribosome protected mRNA fragments (RPFs). In the case of *Selenop*, we surmised that the density of upstream RPFs relative to those downstream of each UGA codon would reflect the efficiency of Sec incorporation. To confirm that the RPFs were derived by protection of mRNA from RNase1 digestion by actively translating ribosomes (i.e. ribosome footprints) we first assessed RPFs aligned to all RefSeq mRNAs. RPFs were approximately 30 nts in size, with 5′ ends that started abruptly 12–13 nts upstream and ended 15–16 nts upstream of annotated start and stop codons, respectively, and were positioned with a strong triplet phasing corresponding to the expected step size (3 nts or 1 codon) of actively translating ribosomes (Supplementary Material S4). RPFs therefore have the expected features of mRNA footprints protected by actively translating ribosomes.

From the ribosome profiling reads mapped to all RefSeq mRNAs, it is possible to predict the codon positions within the RPFs that occupied the ribosomal P- and A-sites. During initiation, the start codon is located in the P-site of the ribosome, whereas during termination the stop codon is in the A-site. Based on where the 5′ ends of the RPFs abruptly start, it can be predicted where the first nt of the P-site and the first nt of the A-site are located. For RPFs less than 31 nts in size, the P-site start codons begin 12 nts downstream from the 5′ end, whereas for RPFs 31 nts or greater in length, the P-sites begin 13 nts downstream from the 5′ end, respectively. To determine if changes

in SELENOP synthesis might alter available selenium pools and affect translation of other selenoproteins, RPFs mapping to regions upstream (CDS5) or downstream (CDS3) of UGA Sec codons were normalized for gene length and total mapped sequence reads in each sample (RPFKMs, RPF reads per kilobase per million mapped reads). Selenoproteins with UGA codons located near the initiation or termination codons were excluded from CDS5 and CDS3 analysis, and *Selenop* is examined separately below. Although a trend for a slight increase in selenoprotein translation was observed in both mutants, none of the changes reached statistical significance (Supplementary Material S5).

To examine the effect of deleting SECIS1 and SECIS2 on translation of *Selenop,* the A-site positions for each RPF were determined along the length of the mRNA, summed at each position and normalized against the total number of mapped RPF reads (reads per million mapped reads) and *Selenop* mRNA abundance (reads per million mapped reads). The results of the SECIS1 deletion mutant and matched wild type controls are shown as blue bars in Figures 6A and 7B, respectively. Analysis of the average A-site count per nucleotide between the start codon, the first UGA, the second UGA, and the termination codon indicates that ribosome density is reduced following the first and second UGA in wild type and SECIS1 deletion mice (Figure 6C). Although the close proximity of UGAs precludes clear discrimination of changes in ribosome density downstream of the second UGA, it is clear that RPFs continue to the end of the coding sequence. Upon deletion of SECIS1 ribosome footprints are relatively unchanged or slightly elevated in relation to wild type up to the second UGA, however RPFs downstream of the second UGA are significantly reduced to near background levels (Figure 6C). We conclude that SECIS1 is more important for translation of UGA codons downstream of the first UGA, compared to the first UGA.

The same analysis was conducted for hepatic translation of *Selenop* in the SECIS2 deletion mutant and matched wild type controls (Figure 6D–F). In contrast to the deletion of SECIS1, ribosome footprint density was increased upstream of the first UGA and decreased between the first and second UGA in the SECIS2 deletion mouse compared to the wild type. This result suggests that the primary role of SECIS2 is during redefinition of the first UGA.

Finally, It should be noted that there is a region of RNase1 protection overlapping SECIS1 that can be seen in all samples except those in which SECIS1 has been deleted. This protection was noted and shown to be independent of SECISBP2 in our previous ribosome profiling studies (40). Understanding the importance of this protection to Sec incorporation and whether it is due to ribosome occupancy or other mechanism(s) merits further investigation.

DISCUSSION

The synthesis of SELENOP is unique in its requirement for the redefinition of multiple UGA codons, which results in the production of several proteins derived from the *Selenop* gene at ratios that may be defined by environmental conditions such as selenium availability. The products include a shortened N-terminal isoform terminating at the second UGA codon that presents the first Sec residue in

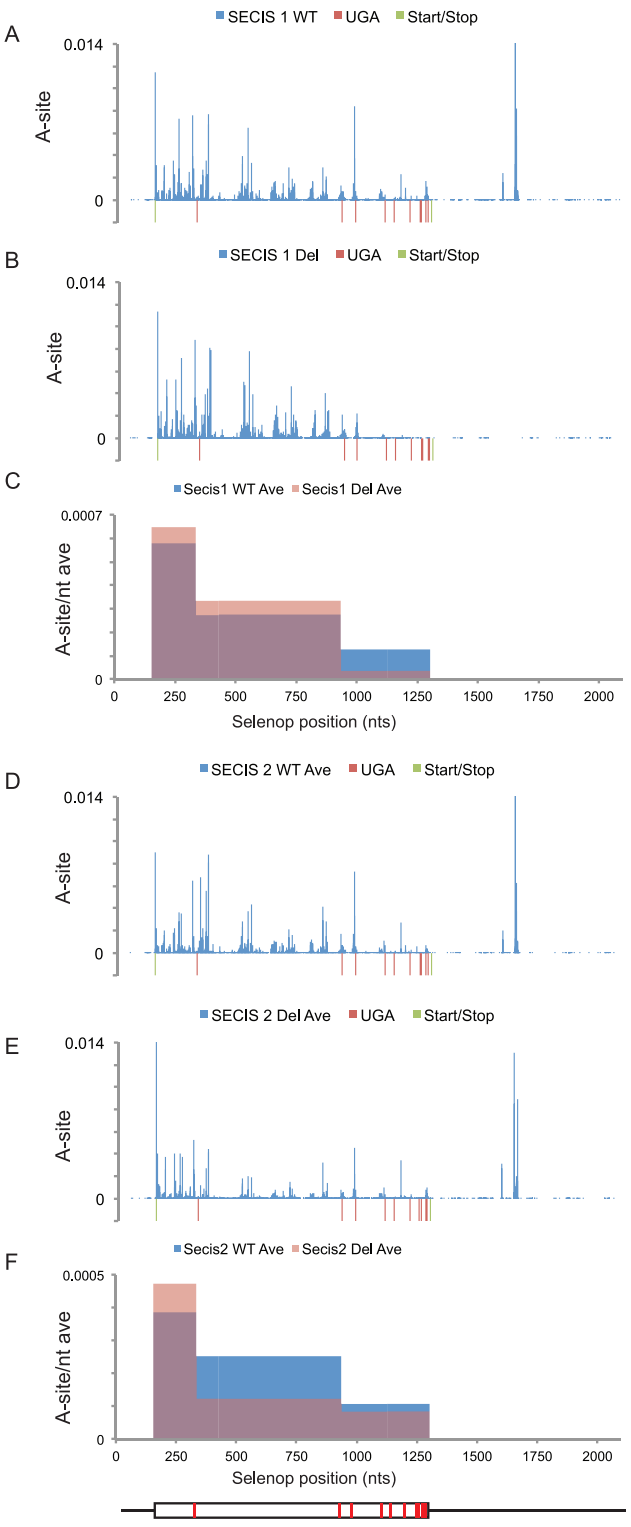

Figure 6. Ribosome profiling of mouse liver in which either SECIS1 or SECIS2 of SELENOP has been deleted. (**A**) Quantifying ribosome footprint A-sites (normalized to RNA levels) across SELENOP in wild type (WT) mouse livers. The position of the start and stop codons (green) as well as UGA codons (red) are indicated below the x-axis. (**B**) Same as A for SECIS1 deleted livers. (**C**) Normalized A-site count per nt across the SELENOP mRNA. Wild type (WT) is in blue, SECIS1 deletion is in orange, overlap is purple. (**D** and **E**) Same as A and B for ribosome profiling of WT and SECIS2 matched livers. (**F**) Same as C for WT (blue) and SECIS2 (orange) samples.

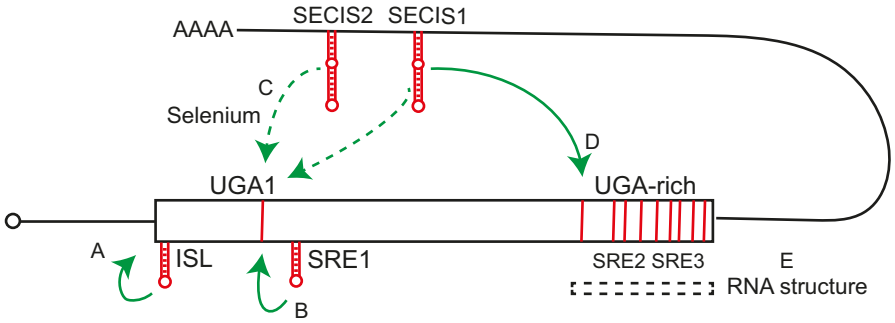

Figure 7. Summary Figure of SELENOP structures and regions and their effects on SELENOP synthesis. (**A**) Indicates the activity of ISL on overall translation. (**B**) Illustrates that SRE1 structure affects UGA redefinition at the first UGA. (**C**) The activity of SECIS2 is sensitive to selenium levels. Dashed green lines represent inefficient redefinition at the first UGA codon by SECIS1 and SECIS2. (**D**) SECIS1 is required for highly efficient UGA redefinition of the UGA codons located near the 3′ UGA-rich end of the coding sequence. (**E**) The UGA-rich region contains SRE2 and SRE3 as well as the potential for other conserved structures, although the higher sequence variability between organisms in this region makes it difficult to define the boundaries relative to UGA codons.

a thioredoxin-like motif with presumed peroxidase activity (24). Longer isoforms that terminate at Sec positions within the C-terminal Sec-rich domain or the natural termination codon are critical for selenium delivery to the brain, testes, and other tissues (27). In addition to the biological importance of SELENOP, this extreme case of codon redefinition raises important questions regarding standard translational decoding and how individual codons can be redefined in a regulated way in different regions of the same mRNA.

Like for virtually all other known cases of translational recoding (2), *cis*-acting RNA structures are fundamental for Sec redefinition. SECIS elements are known to be essential determinants for this process, although the precise roles of each of the two elements present in *Selenop* and their link to Sec incorporation processivity remain to be fully elucidated. Many selenoproteins genes were found to possess additional family-specific RNA structures (e.g. SREs). The functions of these *cis*-acting elements are unclear, but are clearly linked to the efficiency or regulation of UGA redefinition.

In this work, we set out to identify and characterize the function of the RNA structures in vertebrate *Selenop*, combining computational analyses with functional analyses of RNA structures using innovative *in vitro* and cell based reporter assays as well as applying ribosome profiling to quantify Sec incorporation in mice deleted for SECIS1 or SECIS2.

Our comparative search for RNA structures identified four discrete regions predicted to contain nucleotide pairings. These sites consist of the known SECIS elements located in the 3′ UTR, and two new elements termed the ISL and SRE1 that reside near the initiation codon (overlapping the signal peptide sequence) and just downstream of the first UGA, respectively. In addition, a more diffuse pattern of predicted structure was observed in the region encoding the C-terminal Sec-rich region that encompassed putative structures SRE2 and SRE3. The boundaries of the structures in the C-terminal region varied to some degree among species, which is consistent with variable UGA locations and the possibility that the overall structure of the region may be the important feature rather than the linking of individual structures to specific UGA codons. In support of a large overall structure, we found that N- and C- terminal coding sequences of SELENOP were independently capable of processive Sec incorporation *in vitro*, but that both halves of the sequence were required for full efficiency.

Previous observations that several selenoproteins have hypermethylated cap structures at their 5′ end (41) and that initiation from an internal ribosome entry site results in significant reductions in Sec incorporation efficiency (42) suggest that events occurring during translation initiation may be relevant to UGA redefinition events that occur downstream. We observed that mutations of the ISL that overlap the signal peptide sequence of *Selenop* significantly reduced translation efficiency and that this effect was independent of the presence of the first UGA or the inclusion of SECISBP2. It should be noted that recent studies have indicated that RNA elements overlapping the signal peptide-coding region may affect mRNA export from the nucleus (43) as well as translational efficiency (44). As the substitutions designed to disrupt the ISL stem loop also affect both the primary nucleotide sequence and codon usage, we cannot rule out additional effects of the ISL *in vivo*. In an attempt to address these issues, we transfected plasmids expressing the ISL-*Selenop*-reporter fusions into cultured cells. The variability in intracellular luciferase activity and lack of efficient export into the media for both wild type and mutant constructs suggested that the reporter constructs as designed do not reflect normal SELENOP expression, processing, or transport. Nevertheless, we conclude that, when expressed in partially purified reticulocyte translation reactions, the ISL acts to enhance translation initiation on *Selenop* mRNA and that it does not have an effect on redefinition of the first UGA.

Given the previous model proposed by Stoycheva and Berry (30) regarding selective use of SECIS2 at the first UGA codon and SECIS1 at downstream UGAs, we tested the effect of mutating SRE1 downstream of the first UGA with different configurations of 3′ UTR SECIS elements using the newly developed StopGo dual luciferase reporter vector. While mutating SRE1 resulted in a reduction in Sec incorporation efficiency, the magnitude of effect was not dependent upon the identity of the SECIS element or selenium levels. This supports that while SRE1 stimulates Sec

incorporation, it is not involved in SECIS discrimination. This experiment further revealed that constructs containing only SECIS2 were highly sensitive to lowered selenium levels showing an approximately 2-fold decrease in Sec incorporation efficiency when selenium levels were reduced relative to higher selenium levels or when compared to any of the constructs containing SECIS1. This result implies that SECIS2 under certain conditions may act as a sensor of available selenium pools to regulate SELENOP synthesis. Furthermore, our ribosome profiling experiments illustrate that deletion of SECIS2, while not strictly required for Sec incorporation, primarily affects decoding efficiency of the first UGA and that deletion of SECIS1 primarily affects decoding of the downstream UGAs as previously predicted (30).

These results reveal a number of RNA elements acting to control translation, RNA stability, SELENOP secretion and codon redefinition during SELENOP synthesis. In Figure 7, we present a model summarizing our current understanding of how these factors act to facilitate and regulate SELENOP synthesis. Using reporter assays, we find a highly conserved ISL stem loop structure overlapping the signal peptide region that can facilitate translation initiation (A), a SRE1 structure downstream of the first UGA that acts to stimulate Sec incorporation (B), and further we find that while both SECIS1 and SECIS2 can support inefficient Sec incorporation at the first UGA (green dotted lines), the activity of SECIS2 is sensitive to selenium levels (C). Perhaps the latter is due to a selenium triggered riboswitch mechanism or alternatively Sec incorporation efficiency is regulated through indirect effects, such as changes in Sec$tRNA^{[Ser]Sec}$ abundance or modification status, to which SECIS2 is more sensitive. Ribosome profiling of mouse livers carrying a deletion of either SECIS1 or SECIS2 indicates that the loss of SECIS2 primarily affects Sec incorporation at the first UGA, while the loss of SECIS1 affects processive Sec incorporation during translation of the UGA-rich region near the 3′ end of the SELENOP coding sequence (D). In addition to these discrete structures, computational analyses also predict that both the 3′ end of the coding sequence is more structured than the central regions (E), although the exact boundaries of the structure are more difficult to identify due to greater sequence variation in this region between organisms. The factors described here induce the native SELENOP mRNA to adopt distinct structures that act to control translation initiation and codon-specific UGA redefinition efficiencies such that SELENOP synthesis is optimized for cellular needs and selenium delivery throughout the body. We further propose that secondary and tertiary interactions within and possibly between larger regions of the mRNA coding sequence may be involved in coordinating UGA redefinition, rates of translation elongation, and the selective use of SECIS elements by the ribosome during translation of *Selenop*. Future studies will be required to understand how selenium and the overall structure and RNP landscape of *Selenop* contribute to the activity of individual RNA structures described herein.

DATA AVAILABILITY

The RNA-Seq and Ribosome profiling deep sequencing data can be obtained through the National Center for Biotechnology Information, Accession number GSE102890.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Brian Dalley (University of Utah) for assistance with library construction and sequencing of ribosome profiling and RNA-Seq samples and Drs Raymond Burk and Kristina Hill for their assistance with handling of the SECIS1 and SECIS2 deletion mice.

FUNDING

National Institutes of Health [R37 ES02497 to R.B., GM077073 to P.R.C., R01 GM114291 to M.T.H.]; Science Foundation Ireland [13/1A/1835 to J.F.A.]. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. Labunskyy,V.M., Hatfield,D.L. and Gladyshev,V.N. (2014) Selenoproteins: molecular pathways and physiological roles. *Physiol Rev.*, **94**, 739–777.
2. Atkins,J.F. and Gesteland,R.F. (2010) *Recoding: Expansion of Decoding Rules Enriches Gene Expression*. Springer, NY.
3. Berry,M.J., Banu,L., Chen,Y.Y., Mandel,S.J., Kieffer,J.D., Harney,J.W. and Larsen,P.R. (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3′ untranslated region. *Nature*, **353**, 273–276.
4. Berry,M.J., Banu,L., Harney,J.W. and Larsen,P.R. (1993) Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J.*, **12**, 3315–3322.
5. Hill,K.E., Lloyd,R.S. and Burk,R.F. (1993) Conserved nucleotide sequences in the open reading frame and 3′ untranslated region of selenoprotein P mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 537–541.
6. Shen,Q., Chu,F.F. and Newburger,P.E. (1993) Sequences in the 3′-untranslated region of the human cellular glutathione peroxidase gene are necessary and sufficient for selenocysteine incorporation at the UGA codon. *J. Biol. Chem.*, **268**, 11463–11469.
7. Fischer,N., Neumann,P., Bock,L.V., Maracci,C., Wang,Z., Paleskava,A., Konevega,A.L., Schroder,G.F., Grubmuller,H., Ficner,R. *et al.* (2016) The pathway to GTPase activation of elongation factor SelB on the ribosome. *Nature*, **540**, 80–85.
8. Zinoni,F., Heider,J. and Bock,A. (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 4660–4664.
9. Bubenik,J.L., Miniard,A.C. and Driscoll,D.M. (2013) Alternative transcripts and 3′UTR elements govern the incorporation of selenocysteine into selenoprotein S. *PLoS One*, **8**, e62102.
10. Howard,M.T., Aggarwal,G., Anderson,C.B., Khatri,S., Flanigan,K.M. and Atkins,J.F. (2005) Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons. *EMBO J.*, **24**, 1596–1607.
11. Howard,M.T., Moyle,M.W., Aggarwal,G., Carlson,B.A. and Anderson,C.B. (2007) A recoding element that stimulates decoding of UGA codons by Sec tRNA[Ser]Sec. *RNA*, **13**, 912–920.
12. Maiti,B., Arbogast,S., Allamand,V., Moyle,M.W., Anderson,C.B., Richard,P., Guicheney,P., Ferreiro,A., Flanigan,K.M. and Howard,M.T. (2009) A mutation in the SEPN1 selenocysteine redefinition element (SRE) reduces selenocysteine incorporation and leads to SEPN1-related myopathy. *Hum. Mutat.*, **30**, 411–416.

13. Copeland,P.R., Fletcher,J.E., Carlson,B.A., Hatfield,D.L. and Driscoll,D.M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J.*, **19**, 306–314.
14. Copeland,P.R., Stepanik,V.A. and Driscoll,D.M. (2001) Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2. *Mol. Cell Biol.*, **21**, 1491–1498.
15. Fagegaltier,D., Hubert,N., Yamada,K., Mizutani,T., Carbon,P. and Krol,A. (2000) Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.*, **19**, 4796–4805.
16. Tujebajeva,R.M., Copeland,P.R., Xu,X.M., Carlson,B.A., Harney,J.W., Driscoll,D.M., Hatfield,D.L. and Berry,M.J. (2000) Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep.*, **1**, 158–163.
17. Zavacki,A.M., Mansell,J.B., Chung,M., Klimovitsky,B., Harney,J.W. and Berry,M.J. (2003) Coupled tRNA(Sec)-dependent assembly of the selenocysteine decoding apparatus. *Mol. Cell*, **11**, 773–781.
18. Budiman,M.E., Bubenik,J.L., Miniard,A.C., Middleton,L.M., Gerber,C.A., Cash,A. and Driscoll,D.M. (2009) Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. *Mol. Cell*, **35**, 479–489.
19. Chavatte,L., Brown,B.A. and Driscoll,D.M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat. Struct. Mol. Biol.*, **12**, 408–416.
20. Miniard,A.C., Middleton,L.M., Budiman,M.E., Gerber,C.A. and Driscoll,D.M. (2010) Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression. *Nucleic Acids Res.*, **38**, 4807-4820.
21. Gladyshev,V.N., Arner,E.S., Berry,M.J., Brigelius-Flohe,R., Bruford,E.A., Burk,R.F., Carlson,B.A., Castellano,S., Chavatte,L., Conrad,M. *et al.* (2016) Selenoprotein gene nomenclature. *J. Biol. Chem.*, **291**, 24036–24040.
22. Kryukov,G.V. and Gladyshev,V.N. (2000) Selenium metabolism in zebrafish: multiplicity of selenoprotein genes and expression of a protein containing 17 selenocysteine residues. *Genes Cells*, **5**, 1049–1060.
23. Lobanov,A.V., Hatfield,D.L. and Gladyshev,V.N. (2008) Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biol.*, **9**, R62.
24. Kurokawa,S., Eriksson,S., Rose,K.L., Wu,S., Motley,A.K., Hill,S., Winfrey,V.P., McDonald,W.H., Capecchi,M.R., Atkins,J.F. *et al.* (2014) Sepp1(UF) forms are N-terminal selenoprotein P truncations that have peroxidase activity when coupled with thioredoxin reductase-1. *Free Radic. Biol. Med.*, **69**, 67–76.
25. Himeno,S., Chittum,H.S. and Burk,R.F. (1996) Isoforms of selenoprotein P in rat plasma. Evidence for a full-length form and another form that terminates at the second UGA in the open reading frame. *J. Biol. Chem.*, **271**, 15769–15775.
26. Ma,S., Hill,K.E., Caprioli,R.M. and Burk,R.F. (2002) Mass spectrometric characterization of full-length rat selenoprotein P and three isoforms shortened at the C terminus. Evidence that three UGA codons in the mRNA open reading frame have alternative functions of specifying selenocysteine insertion or translation termination. *J. Biol. Chem.*, **277**, 12749–12754.
27. Burk,R.F. and Hill,K.E. (2015) Regulation of Selenium Metabolism and Transport. *Annu. Rev. Nutr.*, **35**, 109–134.
28. Fixsen,S.M. and Howard,M.T. (2010) Processive Selenocysteine Incorporation during Synthesis of Eukaryotic Selenoproteins. *J. Mol. Biol.*, **399**, 385–396.
29. Shetty,S.P., Shah,R. and Copeland,P.R. (2014) Regulation of selenocysteine incorporation into the selenium transport protein, selenoprotein P. *J. Biol. Chem.*, **289**, 25317–25326.
30. Stoytcheva,Z., Tujebajeva,R.M., Harney,J.W. and Berry,M.J. (2006) Efficient incorporation of multiple selenocysteines involves an inefficient decoding step serving as a potential translational checkpoint and ribosome bottleneck. *Mol. Cell Biol.*, **26**, 9177–9184.
31. Lin,H.C., Ho,S.C., Chen,Y.Y., Khoo,K.H., Hsu,P.H. and Yen,H.C. (2015) SELENOPROTEINS. CRL2 aids elimination of truncated selenoproteins produced by failed UGA/Sec decoding. *Science*, **349**, 91–95.
32. Wu,S., Mariotti,M., Santesmasses,D., Hill,K.E., Baclaocos,J., Aparicio-Prat,E., Li,S., Mackrill,J., Wu,Y., Howard,M.T. *et al.* (2016) Human selenoprotein P and S variant mRNAs with different numbers of SECIS elements and inferences from mutant mice of the roles of multiple SECIS elements. *Open Biol*, **6**, 160241.
33. Loughran,G., Howard,M.T., Firth,A.E. and Atkins,J.F. (2017) Avoidance of reporter assay distortions from fused dual reporters. *RNA*,**23**, 1285-1289.
34. Gruber,A.R., Findeiss,S., Washietl,S., Hofacker,I.L. and Stadler,P.F. (2010) RNAz 2.0: improved noncoding RNA detection. *Pac. Symp. Biocomput.*, 69–79.
35. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
36. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
37. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
38. Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 2454–2459.
39. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
40. Fradejas-Villar,N., Seeher,S., Anderson,C.B., Doengi,M., Carlson,B.A., Hatfield,D.L., Schweizer,U. and Howard,M.T. (2016) The RNA-binding protein Secisbp2 differentially modulates UGA codon reassignment and RNA decay. *Nucleic Acids Res.*, **45**, 4094-4107.
41. Wurth,L., Gribling-Burrer,A.S., Verheggen,C., Leichter,M., Takeuchi,A., Baudrey,S., Martin,F., Krol,A., Bertrand,E. and Allmang,C. (2014) Hypermethylated-capped selenoprotein mRNAs in mammals. *Nucleic Acids Res.*, **42**, 8663–8677.
42. Donovan,J. and Copeland,P.R. (2010) The efficiency of selenocysteine incorporation is regulated by translation initiation factors. *J. Mol. Biol.*, **400**, 659–664.
43. Palazzo,A.F., Springer,M., Shibata,Y., Lee,C.S., Dias,A.P. and Rapoport,T.A. (2007) The signal sequence coding region promotes nuclear export of mRNA. *PLoS Biol.*, **5**, e322.
44. Mahadevan,K., Zhang,H., Akef,A., Cui,X.A., Gueroussov,S., Cenik,C., Roth,F.P. and Palazzo,A.F. (2013) RanBP2/Nup358 potentiates the translation of a subset of mRNAs encoding secretory proteins. *PLoS Biol.*, **11**, e1001545.